

Bibliographic data Seed= JP8248993

Epoque PN	JP8248993	19960927
Epoque AN	JP19950052557	19950313
Priority	JP19950052557	19950313

• Classifications: IPC: G10L5/04; G10L3/00

• Applicant: MATSUSHITA ELECTRIC IND CO LTD

• Inventors: KAMAI TAKAHIRO; HARA KIYO; MATSUI KENJI

• Titles : T1: CONTROLLING METHOD OF PHONEME TIME LENGTH

• Abstract PURPOSE: To perform an optimum control of a phoneme time length and to conduct a rhythmical and natural synthesis without adversely affecting the features of a phoneme by controlling the time length of the sum of a vowel and a consonant located right after the vowel so that the time length becomes a prescribed one. CONSTITUTION: A rhythm generating section 3 outputs equal interval rhythms in accordance with the average time length of one syllable which is decided beforehand. Then, voice piece sounding timing generation section 4 is operated for the respective phoneme numbers set by a phoneme selecting section 2. If a phoneme number represents a normal CV syllable, the section 4 reads the consonant time of the corresponding phoneme from a consonant time length storage section 7 in accordance with the phoneme number.

The section 4 sets a phoneme sounding timing at the time which is defined by going back the equal interval rhythm received from the section 3 for equivalent of the consonant time length. Accordingly, the boundaries of the consonant and the vowel of each syllable are arranged with equal intervals and the synthesized sound is heard with a regular rhythm.

BEST AVAILABLE COPY

2005/12/09

BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-248993

(43) 公開日 平成8年(1996)9月27日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 5/04			G 1 0 L 5/04	F
				B
3/00			3/00	H

審査請求 未請求 請求項の数9 O L (全 8 頁)

(21) 出願番号 特願平7-52557

(22) 出願日 平成7年(1995)3月13日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 釜井 孝浩

大阪府門真市大字門真1006番地 松下電器産業株式会社内

(72) 発明者 原 紀代

大阪府門真市大字門真1006番地 松下電器産業株式会社内

(72) 発明者 松井 謙二

大阪府門真市大字門真1006番地 松下電器産業株式会社内

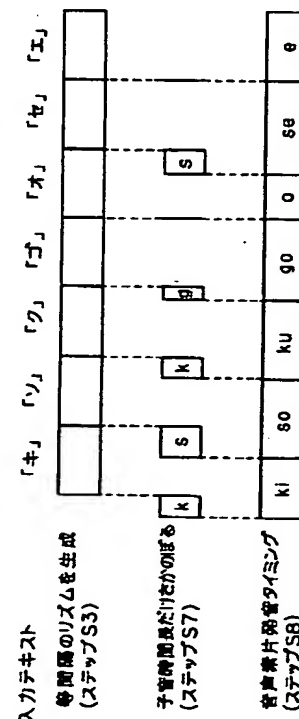
(74) 代理人 弁理士 滝本 智之 (外1名)

(54) 【発明の名称】 音韻時間長制御方法

(57) 【要約】

【目的】 リズミカルで自然な音韻時間長制御を簡易に実現すること。

【構成】 C V 音声素片の子音時間長を記憶しておき、等間隔に生成したリズムから該当するC V 音声素片の子音時間長分さかのぼった位置に、各C V 音声素片の発音開始タイミングを設定する。その結果、C V 音声素片の子音から母音への境界部が等間隔に並ぶ。その結果、リズムミカルで自然な音声合成できる。



【特許請求の範囲】

【請求項 1】母音とその直後の子音を合わせた時間長が所定の時間長になるように制御することを特徴とする音韻時間長制御方法。

【請求項 2】所定の時間長は等間隔のリズムであることを特徴とする請求項 1 記載の音韻時間長制御方法。

【請求項 3】子音とその直後の母音の対である CV 音節を単位とし、その単位で作成された CV 音声素片を複数接続して音声を作成する、CV 単位素片接続型音声合成方法において、あらかじめ生成したリズムから CV 音声素片上の子音部分の時間長である子音時間長をさかのぼった位置に、CV 音声素片発音開始タイミングを設定することを特徴とする音韻時間長制御方法。

【請求項 4】音節が母音または撥音であるとき、子音時間長を 0 と見なすことを特徴とする請求項 3 記載の音韻時間長制御方法。

【請求項 5】音節が促音であるとき、あらかじめ決めておいた仮想子音時間長を子音時間長と見なすことを特徴とする請求項 3 または請求項 4 記載の音韻時間長制御方法。

【請求項 6】音節が無声化音であるとき、子音時間長は無声化音節の音声素片の子音時間長であり、無声化でなかった場合に本来存在すべき母音部分の時間長を、その音節の次の音節の種類に応じて処理することを特徴とする請求項 3 から請求項 5 のいずれかに記載の音韻時間長制御方法。

【請求項 7】次の音節が「サ」行または「シャ」行またはそれに類する音節の場合、本来存在すべき母音部分の時間長は 0 とし、代わりに次の音節の先頭の子音部分の時間長を長くすることによって時間長を補償することを特徴とする請求項 6 記載の音韻時間長制御方法。

【請求項 8】次の音節が「サ」行ではなく、かつ「シャ」行ではなく、かつ「サ」行に類する音節ではなく、かつ「シャ」行に類する音節ではない場合、本来存在すべき母音部分の時間長は 0 とし、代わりに次の音節の先頭部分に無音部分を追加することによって時間長を補償することを特徴とする請求項 6 又は請求項 7 記載の音韻時間長制御方法。

【請求項 9】あらかじめ生成したリズムは等間隔のリズムであることを特徴とする請求項 3 から請求項 8 のいずれかに記載の音韻時間長制御方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、自動読書器、電子メール読み上げ装置、電話応答システムなどに応用可能なテキスト／音声変換技術、すなわち音声規則合成技術に関するものである。

【0002】

【従来の技術】音声規則合成技術は任意のテキストを入力とし、音声に変換して出力するもので、表示部を持た

ない機器や、カーナビゲーションシステムのように人間が視覚を使えない状況で使用される機器などで、機器から人間への情報手段に広く用いることのできる技術である。また、マルチメディア技術としてもテキスト化された情報を人間にわかりやすい形式で伝達する手段として有効である。

【0003】このようないわばインタフェース技術としての音声規則合成技術には、合成音声の明瞭度が高いことも重要であるが、音声合成にともなう処理が単純であることが重要であることも多い。なぜなら、例えばパソコンなどのマルチメディアの枠組みで動作させる場合には CPU 資源を音声出力以外の処理に多く占有されることが多く、結果的に限られた処理能力で音声出力を可能としないといけないからである。

【0004】また、電子ブックなどの携帯型の装置に組み込むことは、音声出力を可能にすると同時に液晶表示部を簡素化するなどのメリットがあるが、そのような小型で安価な装置には高性能な CPU が使用できないことが多い。

【0005】以上のような観点から、音声規則合成技術には処理量が少ないことが重要であると言える。

【0006】そのような目的で音声素片編集型の規則合成方法が発明されている（特願平 6-302471）。そのシステムでは音節単位の音声波形素片を編集することにより任意の文章を合成する。音声波形素片を用いることで、例えばフォルマント合成などのパラメトリック合成系におけるフィルタ演算が不要になり、処理量が低減できる。

【0007】しかし、なめらかな音声を合成するには用意された音声素片を目的に応じて変形する操作が必要である。特に重要な変形はピッチと時間長である。

【0008】ピッチの変形は図 5 のように行われる。すなわち、あらかじめピッチ周期に対応したピークを中心として窓関数によりピッチ波形を切り出しておき、合成時に目的のピッチ周期の間隔で重ね合わせることでピッチを変更する。

【0009】時間長の変更は音声素片の定常母音部分で行う。時間長を短くするときは母音部の途中で波形再生を打ち切り、長くするときは同じピッチ波形を繰り返す。子音の時間長変更については、発話速度の変化に対して子音時間長の変化が小さいことと、子音の種類により異なる時間長変更操作が必要なことから、実施しない。

【0010】さて、そのような枠組みのもとで自然な発話に聞こえるように時間長制御を行うには、従来子音時間長と母音時間長を統計から決定し、音節の時間長を子音時間長と母音時間長の和として決めていた。例えば、音節「か」の時間長 D は、統計から求められた子音 / k / の時間長 D_k と母音 / a / の時間長 D_a の和として（数 1）により決定する。

【0011】

$D = D_k + D_a \quad \dots \dots \quad (\text{数}1)$

しかし、こうして統計から求められた時間長は、必ずしも音声素片固有の時間長とマッチしない。そのため、音節内の子音と母音の時間長のバランスが悪くなる。

【0012】このような問題は、時間長の決定を音声素片の持つ時間長と関わりなく行うことによって起こる。例えば音節時間長がほぼ均一になるように時間長を決定しても、等間隔のリズムで聞こえないということが起こる。

【0013】

【発明が解決しようとする課題】以上説明したように従来の音韻時間長決定方法は、音声素片を用いた合成方法において音声素片固有の音韻時間長との不整合を起こし、聞こえに対して最適な音韻時間長の制御が行えないという問題があった。

【0014】

【課題を解決するための手段】上記の問題を解決するために、本発明の音韻時間長制御方法は、あらかじめ音声素片の子音と母音の境界にマークを付与しておき、その子音の時間長を変更せずにそのまま合成時の子音時間長として用いる。その動作は、まず等間隔のリズムを生成し、その各タイミングから子音時間長だけさかのぼったタイミングを音声素片発音タイミングとする。各音声素片の長さの制御は母音部分の時間長を変更することで行い、子音の時間長は変更しない。

【0015】

【作用】上記の手段により、音声素片の子音と母音の境界が等間隔に並ぶようになる。すなわち、合成音声の母音開始部から子音終了部までを一つの単位（VC単位）とすると、VC単位が一定の時間長になるように作用する。

【0016】

【実施例】次に、本発明の音韻時間長制御方法について、図面を参照しながら説明する。

【0017】図1に本発明の音韻時間長制御方法の構成を示す。その構成では、入力部1が設けられ、その出力は音声素片選択部2に接続されている。また、リズム生成部3が設けられ、その出力は音声素片発音タイミング生成部4に接続されている。また、音声素片記憶部8と音声素片読みだし部5が設けられ、それらは互いに接続されている。音声素片発音タイミング生成部4の出力は音声素片読みだし部5に接続されている。そして、子音時間長記憶部7が設けられ、音声素片発音タイミング生成部4に接続されている。また、音声素片選択部2の出力は音声素片読みだし部5と音声素片発音タイミング生成部4に接続されている。音声素片読みだし部5の出力は出力部6に接続されている。

【0018】続いて、本構成による音韻時間長制御方法の手順をフローチャートを参照しながら説明する。図2

は本発明の全体の処理の流れを表したフローチャートである。

【0019】まず、音声素片記憶部8には音声合成に必要な全ての音声波形が一つの子音とそれに続く母音の対である音節を単位として記憶されている。また、子音時間長記憶部7には音声素片記憶部8に記憶された全ての音声素片の子音時間長が記憶されている。

【0020】入力部1は1行分の発音記号列を読み込み、音声素片選択部2に出力する（ステップS1）。音声素片選択部2は発音記号に対応する音声素片を表す音声素片番号を出力する（ステップS2）。また、リズム生成部3はあらかじめ決められた1音節の平均時間長に従って、等間隔のリズムを出力する（ステップS3）。

【0021】続いて音声素片選択部2によって設定された音声素片番号それぞれについて、音声素片発音タイミング生成部4が動作する。

【0022】音声素片番号が通常のCV音節を表すとき（ステップS4）、音声素片発音タイミング生成部4は音声素片番号に従って子音時間長記憶部7から、対応する音声素片の子音時間長を読み出す（ステップS5）。ここで、通常のCV音節とは子音と母音の一つずつの組み合わせからなる音節を指し、母音、撥音（「ン」）、促音（「ッ」）、無声化音節以外の物を言う。

【0023】音声素片番号が通常のCV音節以外の場合には例外処理を行う（ステップS6）。例外処理については後述する。

【0024】音声素片発音タイミング生成部4はリズム生成部3から受け取った等間隔のリズムから子音時間長分さかのぼり（ステップS7）、その時刻に音声素片発音タイミングを設定する（ステップS8）。このようにすることで、図3に示すように各音節の子音と母音の境界が等間隔に並ぶようになる。このことで、合成音が規則的なリズムで聞こえる効果がある。

【0025】音声素片読みだし部5は音声素片記憶部8から音声素片番号に対応する音声素片を、音声素片発音タイミング生成部4から受け取った音声素片発音タイミングに従って読み出し、出力部6に対して出力する（ステップS9）。出力部6は音声素片読みだし部5から受け取った音声素片波形を音声に変換して出力する。

【0026】次に例外処理についてフローチャートを参照しながら説明する。図4は例外処理の流れを表すフローチャートである。

【0027】まず、母音と撥音の場合（ステップS101）、子音時間長は0とし（ステップS102）、通常と同様の処理を行う。すなわち、等間隔のリズムをそのまま音声素片発音タイミングとみなす。

【0028】音声素片番号が促音の場合（ステップS103）、その音節は仮想的に子音と母音に分かれているとみなし、あらかじめ決めておいた仮想子音時間長（例えば40ms）を子音時間長として用い（ステップS1

04)、以下、通常の場合と同様の処理を行う。

【0029】音声素片番号が無声化音節の場合(ステップS105)、本来存在した母音がなくなった物と考え、音声素片全体の長さを子音時間長とみなし(ステップS106)、通常と同様に等間隔のリズムからさかのぼって音声素片発音タイミングを生成する。しかし、次に来るべき母音が存在しないため、次の音節までの間に無音区間が生まれる。そこで、次の音節の種類により無音区間の処理を行う。

【0030】無声化音節の次に来る音節は通常の場合無声子音のみである。まず、その無声子音が「カ」行、「タ」行、「バ」行の場合、無音区間はそのまま無音とする。これらの子音は閉鎖子音であり、先頭部分にはもともと特有の無音区間が存在するので、上の処理は元々存在した無音区間を長くするよう作用する。無音区間が長くなっても音韻性(その音韻らしく聞こえる度合い)が損なわれることはないし、等間隔のリズムが保証されているので、促音に異聴することもない。

【0031】無声化音節の次に来る無声子音が「ハ」行の場合も同様にする。「ハ」行は閉鎖子音ではないが、先頭に無音区間が付加されても異聴などは起こらない。

【0032】無声化音節の次の無声子音が「サ」行の場合は別の処理を行う(ステップS107)。もし上と同様の処理を行うと音韻性の低下を引き起こす。すなわち、無音区間が付加されることにより「ツァ」行への異聴が発生する。そこで、無音区間がなくなるように

「サ」行の子音部分を長くする(ステップS108)。このことは、子音母音境界が等間隔のリズムになることを保証した上で、無声化により無音になった区間を「サ」行の子音の波形で埋めるように作用する。

【0033】「サ」行の子音時間長を長くする方法は例えば次のようにする。音声素片を録音するとき、子音部分が長めになるようにする。通常、単独発声で「サ」行の単音節を発声すると、必要な子音時間長よりも十分長い子音時間長となる。そこで、普段は子音部分の読みだしを途中から行うことで子音時間長を短くしておき、長くする操作が必要なときは読みだし開始位置を早める。この方法は処理が単純であり、効果的である。

【0034】その他の無声子音としては「キャ」行、「チャ」行、「ピャ」行、「ヒャ」行、「シャ」行があるが、それぞれ「カ」行、「タ」行、「バ」行、「ハ」行、「サ」行と同じ処理を行う。

【0035】このような時間長制御を行うことにより、音声素片の子音から母音への境界が等間隔になり、規則正しいリズムの合成音が生成できる。日本語の音声は等間隔のリズムが特徴とされているが、音節中のどのようなタイミングが等間隔になっているかは明らかになっていない。本方法では子音から母音への変化点が等間隔になるが、実際そのように制御することで等間隔のリズムで発話しているように聞こえる。

【0036】本方法によれば、複雑な時間長の計算も不要である。さらに、音声素片の時間長変更処理は処理が単純な母音定常部で行うことが基本であり、子音の時間長変更は一部の子音を除いて不要である。「サ」行、「シャ」行などのように時間長変更が必要な子音も、その読みだし開始位置を変更するだけでよいので、複雑な処理は不要である。

【0037】また、歌唱の合成の際には音楽のリズムに合わせて正確な発音タイミングを生成しなくてはならない。このような場合、子音から母音への境界がリズムに合うように制御することで、自然な歌唱が合成できる。

【0038】なお、本方法では音声素片発音タイミング生成のための基本リズムとして、等間隔のリズムを用いたが、これは等間隔でなくても構わない。例えば、文頭から文末にかけて徐々に長くなるようなリズムを用いることも考えられるし、統計データを用いて音韻の種類ごとに時間長を設定してももちろん構わない。本方法を用いれば、子音から母音への変化タイミングが尊重されるので、そのような統計データがあれば直ちに利用できる。すなわち、自然発話のデータを用いて母音開始部から子音終了部までの時間の統計を音韻の組み合わせごとに調べておき、合成時に利用することができる。

【0039】また、ここでは音声素片を用いた合成を例にとり説明したが、子音から母音への境界を等間隔にするなどの制御をフォルマント合成などのパラメトリックな合成方式に用いても、もちろん構わない。

【0040】

【発明の効果】以上説明したように、本発明による音韻時間長制御方法を用いれば、子音から母音への境界が、等間隔のリズムなどの意図したタイミングに一致するように制御できるため、リズムカルで自然な合成が行える。また、処理の難しい子音部分の時間長変更をほとんど行う必要がなく、音声素片の持つ特徴を損なわずに高品質な合成が可能になる。

【図面の簡単な説明】

【図1】本発明の音韻時間長制御方法の構成図

【図2】本発明の音韻時間長制御方法の全体の処理の流れを表すフローチャート

【図3】本発明の音韻時間長制御方法の各段階における時間長データの変化の様子の説明図

【図4】本発明の音韻時間長制御方法の例外処理の流れを表すフローチャート

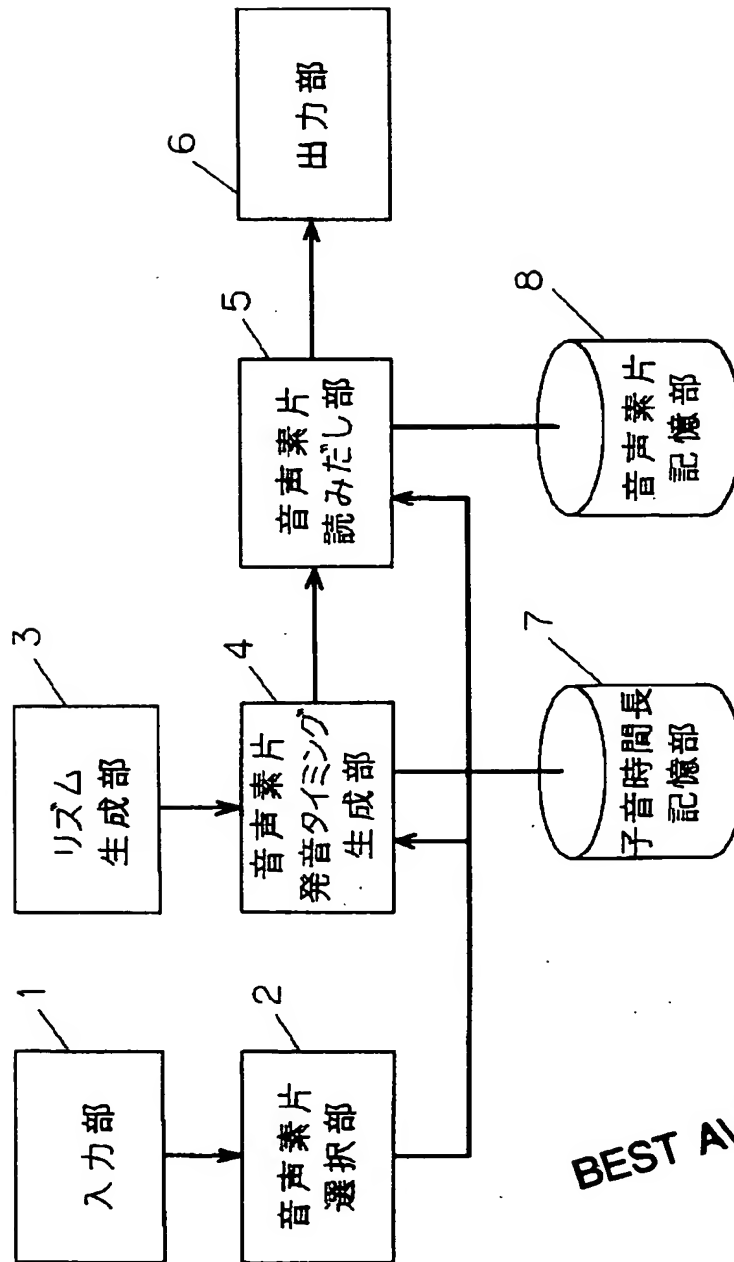
【図5】音声素片のピッチを変更する様子を表す説明図
【符号の説明】

- 1 入力部
- 2 音声素片選択部
- 3 リズム生成部
- 4 音声素片発音タイミング生成部
- 5 音声素片読みだし部
- 6 出力部

7 子音時間長記憶部

8 音声素片記憶部

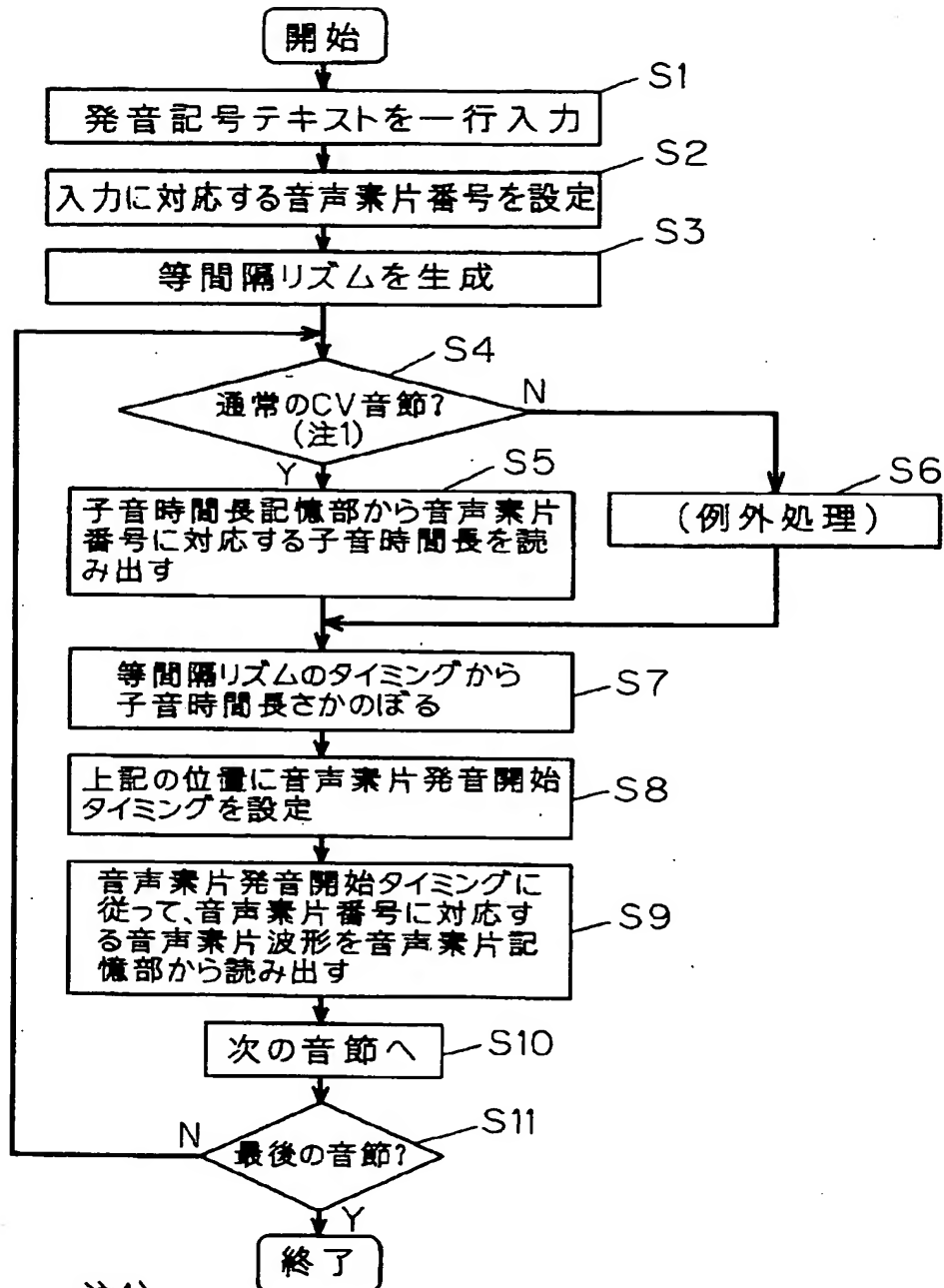
【図1】



BEST AVAILABLE COPY

BEST AVAILABLE COPY

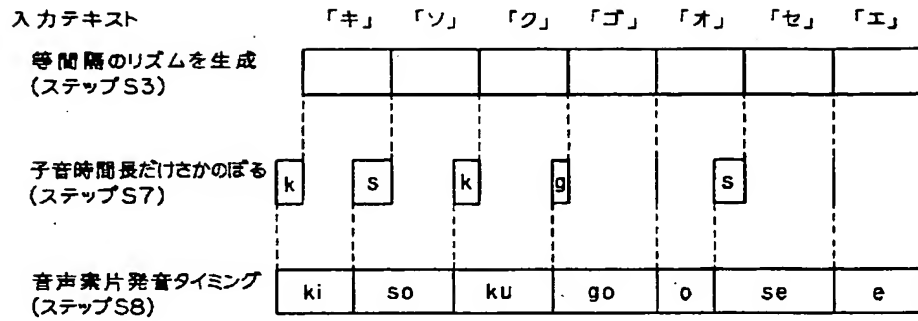
【図2】



注1)

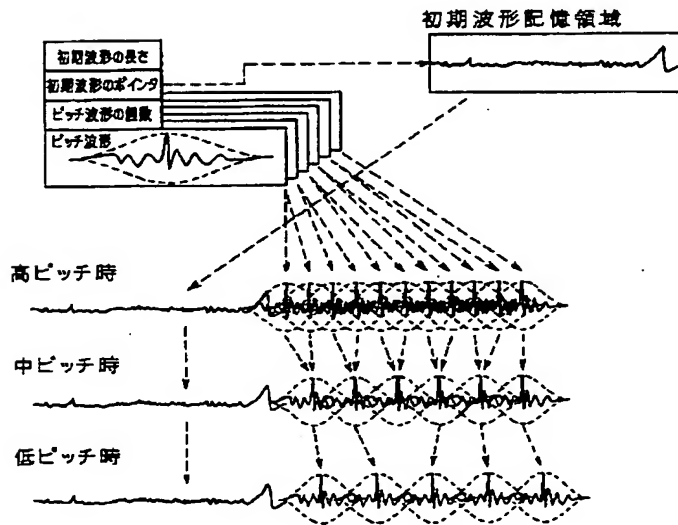
ここでは母音、撥音、促音、無声化音節以外の
子音と母音の一つずつの組み合わせによる
音節を通常のCV音節と言う。

【図3】



BEST AVAILABLE COPY

【図5】



BEST AVAILABLE COPY

【図4】

